Facing Uncertainty in Data Analysis: Ensuring Data and Information Comparability

Bill Ingersoll

Navy Quality and Accreditation Office

843-764-7337

DSN: 794-7337

ingersollws@navsea.navy.mil

www.navylabs.navy.mil

Discussion Topics

◆ Environmental Decision Quality

◆ Data Quality Comparability Issues

◆ Estimating Uncertainty for Quantitative Data

   ◆ All Quantitative Measurements Must Have an Associated Estimation of Uncertainty

◆ Managing Decision Uncertainty for Qualitative Data

   ◆ Estimated and Censored Data

Environmental Decision Quality

◆ To Make the Right Environmental Decisions Requires Understanding the Quality of the Data

◆ Data Comparability is an Important Component in Data Quality

◆ Ensuring Data Comparability Requires Estimating and Minimizing Analytical Measurement Uncertainty

Data Comparability Issues

◆ Federal and State Spend > $1 Billion/year to Monitor Water Quality

◆ Problems with Data Quality Consistency

◆ Difficult to Share Data Between Agencies Because of Data Quality Issues

◆ Information About the Data Quality is Not Readily Available

Advantages of Data Quality Comparability

◆ Integration of Data From Different Study Sources

◆ Collection of Data of Know Quality

◆ Collaborative Monitoring Information for Decision Making

Example of Data Quality Comparability

- Comparability of Data
  - Study A Result: 10 mg/L
  - Study B Result: 10 mg/L
  - Study C Result: $\leq$ 20 mg/L

Example of Data Quality Comparability

- Comparability Without Estimated Uncertainty
  - Study A Result: 10 mg/L
  - Study B Result: 10 mg/L
  - Study C Result: $\leq$ 20 mg/L

- Comparability Using Estimated Uncertainty
  - Study A Result: 10 +/- 2 mg/L
  - Study B Result: 10 +/- 10 mg/L
  - Study C Result: $\leq$ 20 mg/L

What Data Are Comparable in Quality?

◆ Comparability Without Estimated Uncertainty

  - ◆ Study A result: 10 mg/L
  - ◆ Study B result: 10 mg/L
  - ◆ Study C result: $\leq$ 20 mg/L

◆ Comparability Using Estimated Uncertainty

  - ◆ Study A result: 10 +/- 2 mg/L
  - ◆ Study B result: 10 +/- 10 mg/L
  - ◆ Study C result: $\leq$ 20 mg/L

Data Quality ⇓ Decision Error ⇑

◆ Water Quality Criteria = 100 ppb

◆ Measurement = 70 ppb

◆ Will the Correct Decision be Made?

◆ Yes, If $X = 70 \pm 20$ ppb (50 – 90 ppb)

◆ Maybe Not, If $X = 70 \pm 40$ ppb (30 – 110 ppb)

Accounting for Data Uncertainty in Decision Making

$$\left[Measurement \pm \frac{t \times \sigma_{Total\_Study}}{N^{1/2}} \right]$$

$$\sigma^2_{Total\_Study} = \sigma^2_{Site} + \sigma^2_{Sampling} + \sigma^2_{Testing}$$

Conceptual Model For Estimating Uncertainty

Total Study Uncertainty, ^{T}S

Site Variability, ^{S}S
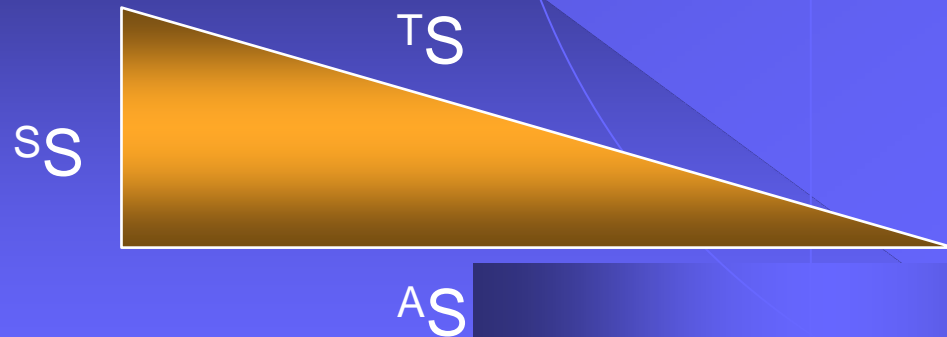
Sampling and Testing Variability, ^{A}S

$^{T}S^2 = {}^{S}S^2 + {}^{A}S^2$

$^{T}S^2 = 30^2 + 10^2$

$^{T}S = 32$

^{T}S

^{S}S

^{A}S

Analytical Measurement Uncertainty

Accurate Laboratory Testing	+	Non-Representative Field Sampling

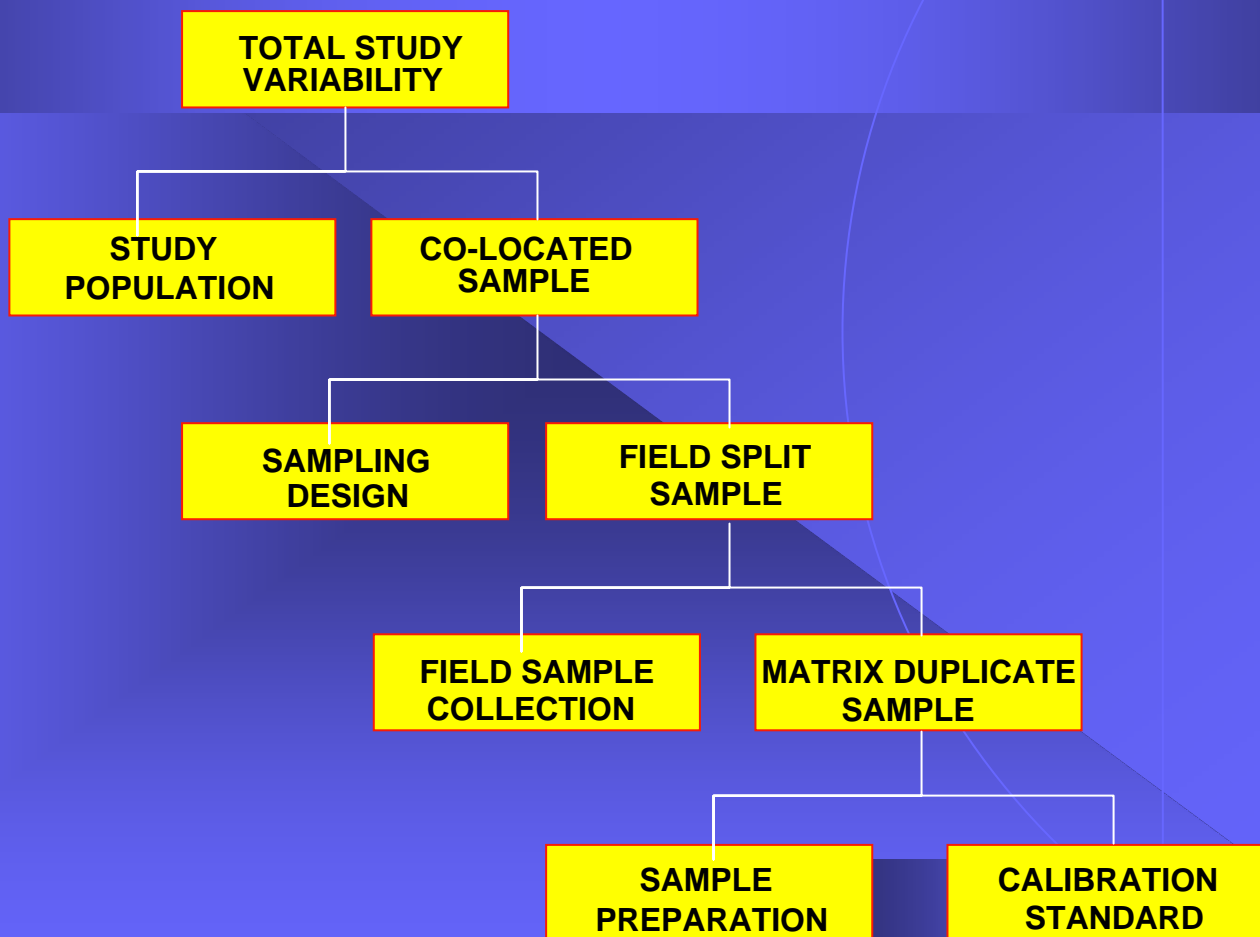$$\Downarrow$$
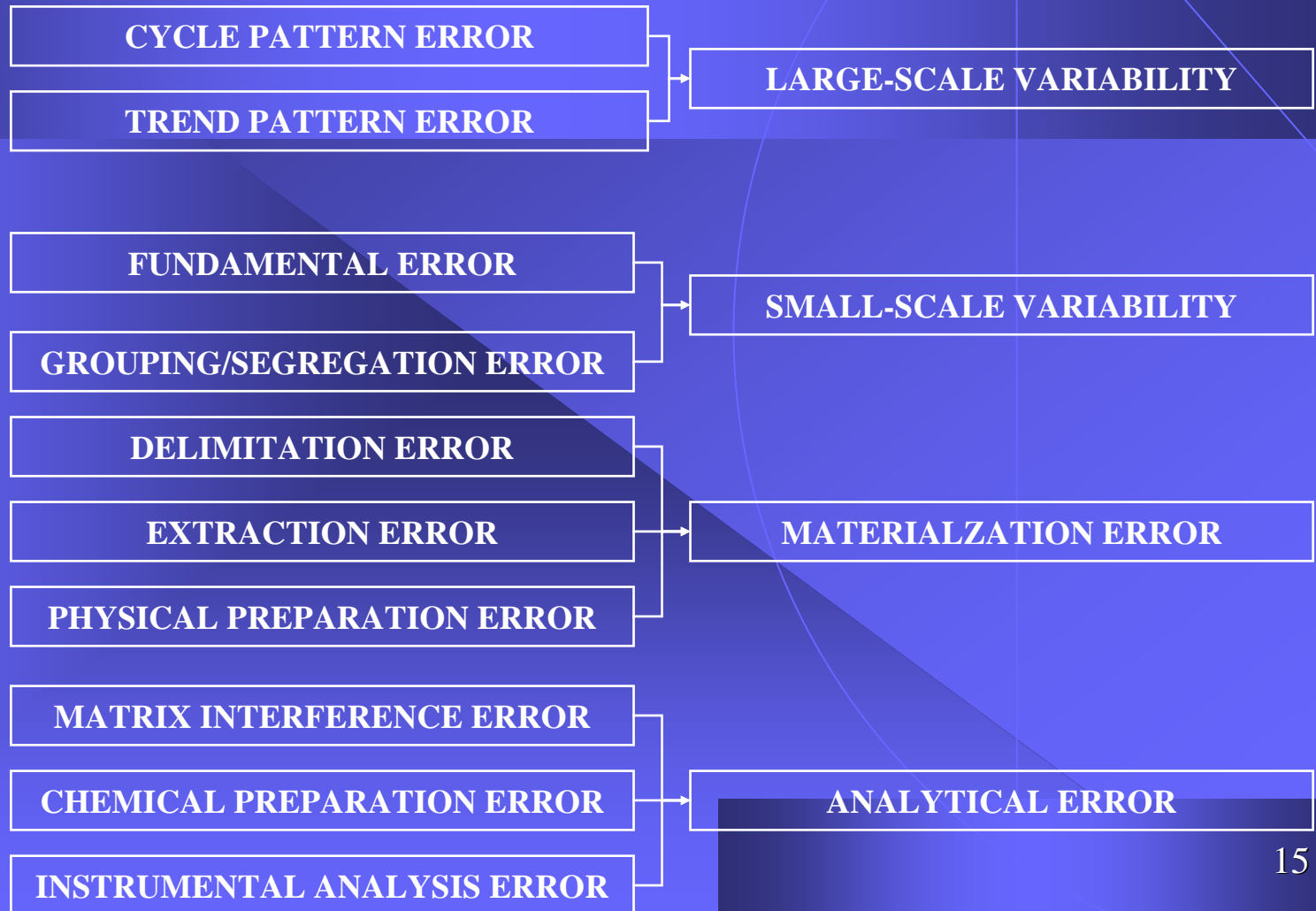
BAD DATA

$$\Downarrow$$

BAD DECISION

Data Comparability Problems and Solutions

- Systemic Failure to Capture the Magnitude of Data Variability

- Generic Data Sets Poorly Matched to Decision-Making Needs

- Distinguish Between Analytical Quality and Data Quality

- Include Uncertainty Estimation in Data Reporting and Uncertainty Management in Decision-Making

Total Study Variability:
Hierarchy of Components

Sampling and Testing Components

CYCLE PATTERN ERROR

TREND PATTERN ERROR

→ **LARGE-SCALE VARIABILITY**

FUNDAMENTAL ERROR

GROUPING/SEGREGATION ERROR

→ **SMALL-SCALE VARIABILITY**

DELIMITATION ERROR

EXTRACTION ERROR

PHYSICAL PREPARATION ERROR

→ **MATERIALZATION ERROR**

MATRIX INTERFERENCE ERROR

CHEMICAL PREPARATION ERROR

INSTRUMENTAL ANALYSIS ERROR

→ **ANALYTICAL ERROR**

15

Variability Factors

Large-scale

Small-scale

Materialization

ANALYTICAL → **MATRIX SAMPLE**

Variability Factors

Large-scale

Small-scale

MATERIALIZATION

ANALYTICAL

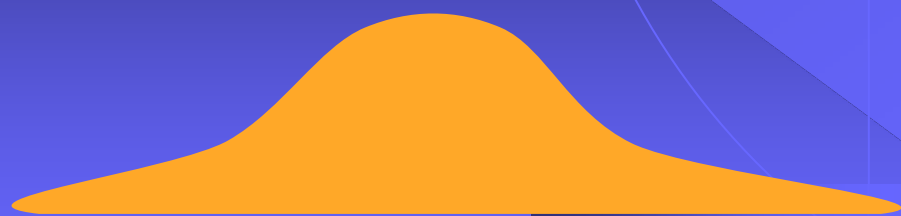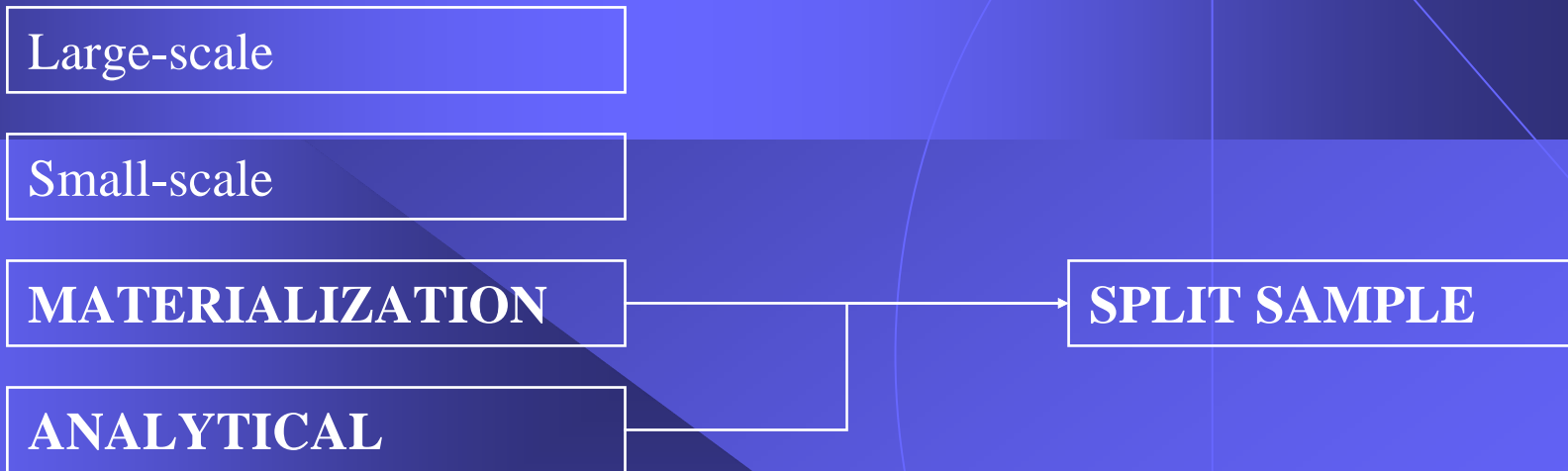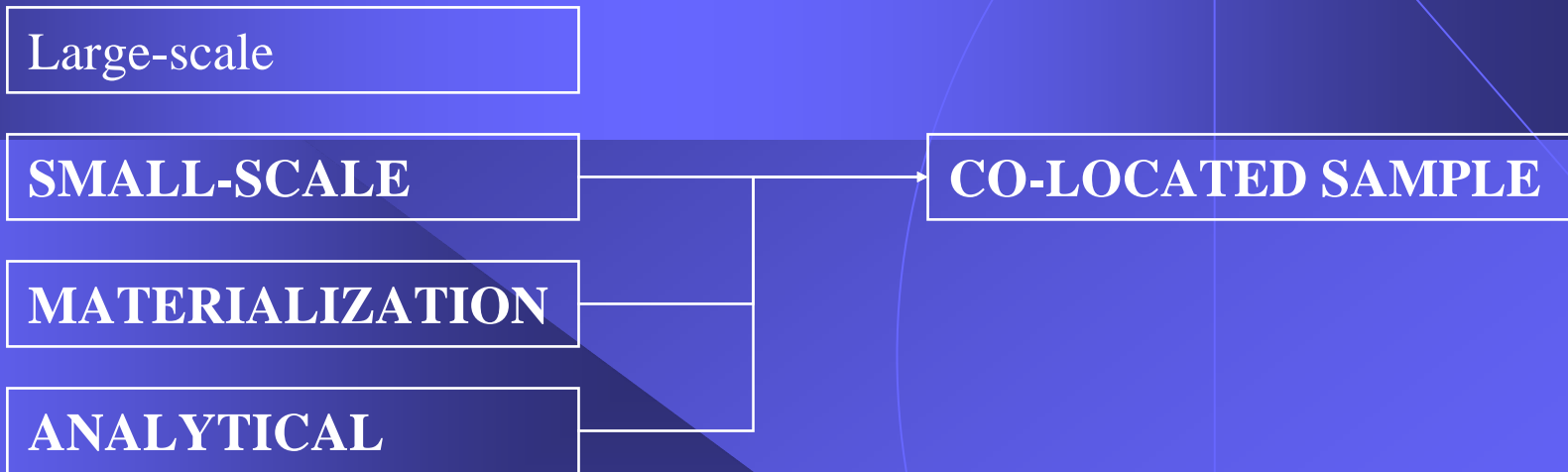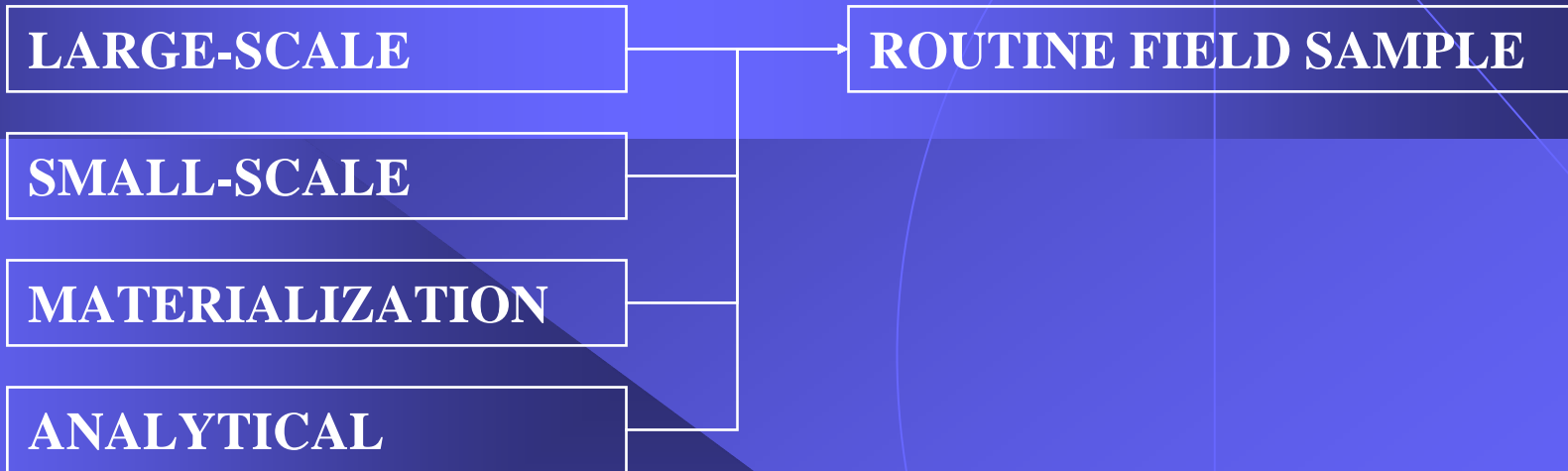**SPLIT SAMPLE**

Variability Factors

Large-scale

SMALL-SCALE → CO-LOCATED SAMPLE

MATERIALIZATION

ANALYTICAL

Variability Factors

| LARGE-SCALE | → | ROUTINE FIELD SAMPLE |

LARGE-SCALE

SMALL-SCALE

MATERIALIZATION

ANALYTICAL

ROUTINE FIELD SAMPLE

Multiple Increment Sampling

Population Distribution

$n = 2$

$n = 4$

$n = 25$
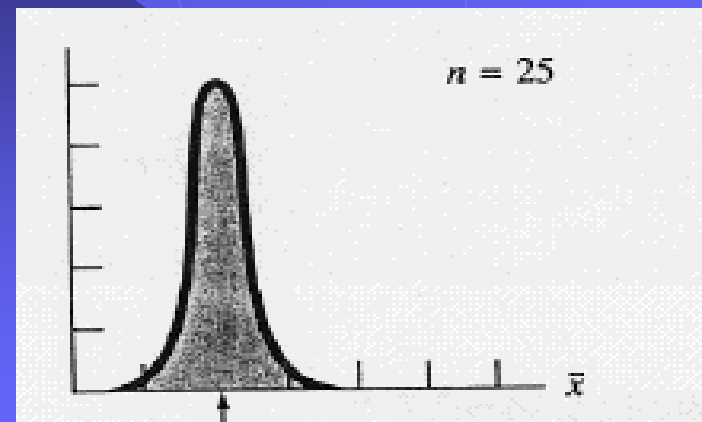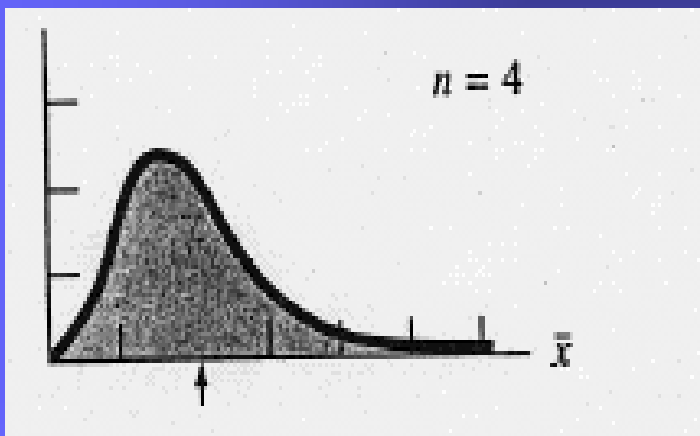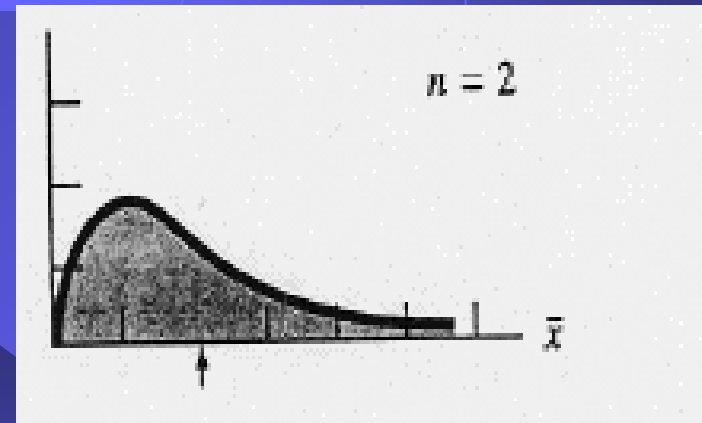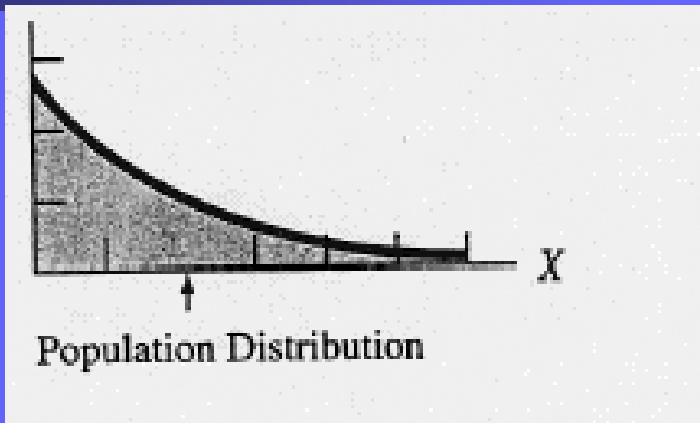
Fully-Nested Hierarchical Design

- Field Sample
- Co-Located
- Field Split
- Preparation
- Test
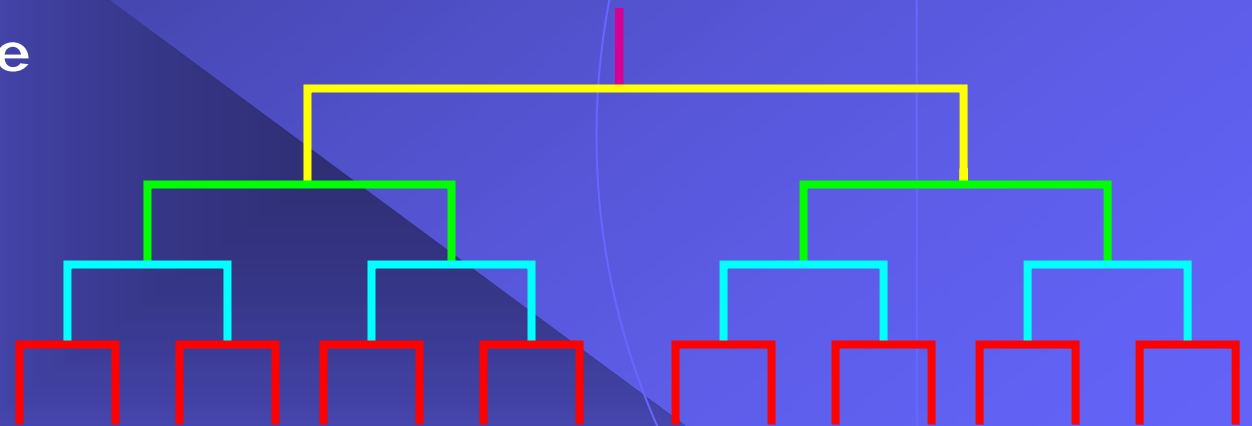
What are the analyte/matrix/technology? | Copper in Wastewater by ICP

Enter 20 replicate results for the following quality control samples as percent deviation (%):

ICS - Instrument calibration standard

**ICV - Second source calibration verification standard**

LCS - Laboratory control sample

MIS - Matrix interference sample (matrix spike, organic surrogate, radiochemical tracer)

FDS - Field-split duplicate sample

CLS - Co-located duplicate sample

	ICS	ICV	LCS	MIS	FDS	CLS
	1.1	0.5	4.0	12.0	0.0	0.0
	0.8	0.1	0.5	1.4	0.0	0.0
	0.4	1.0	1.5	8.0	0.0	0.0
	2.0	1.2	1.7	3.7	0.0	0.0
	1.0	0.2	0.1	12.0	0.0	0.0
	1.2	0.4	2.2	0.4	0.0	0.0
	1.7	1.2	0.4	3.6	0.0	0.0
	3.7	0.9	0.3	0.1	0.0	0.0
	1.1	0.1	0.5	2.7	0.0	0.0
	3.1	1.3	15.0	17.0	0.0	0.0
	2.0	0.9	20.0	30.0	0.0	0.0
	0.7	1.0	0.4	3.7	0.0	0.0
	0.4	2.0	4.0	1.5	0.0	0.0
	0.9	0.2	0.6	5.0	0.0	0.0
	1.4	1.0	1.5	1.4	0.0	0.0
	1.9	1.4	5.0	20.0	0.0	0.0
	2.0	1.5	24.0	3.5	0.0	0.0
	1.5	1.7	3.0	5.0	0.0	0.0
	1.6	3.0	13.0	-24.0	0.0	0.0
	1.1	3.1	11.0	-13.0	0.0	0.0
Std. Dev.	0.84	0.85	7.2	11.1	0.0	0.0
Bias	1.5	1.1	5.4	4.7		
Recovery	101.5	101.1	105.4	104.7		

Partitioning of Uncertainty
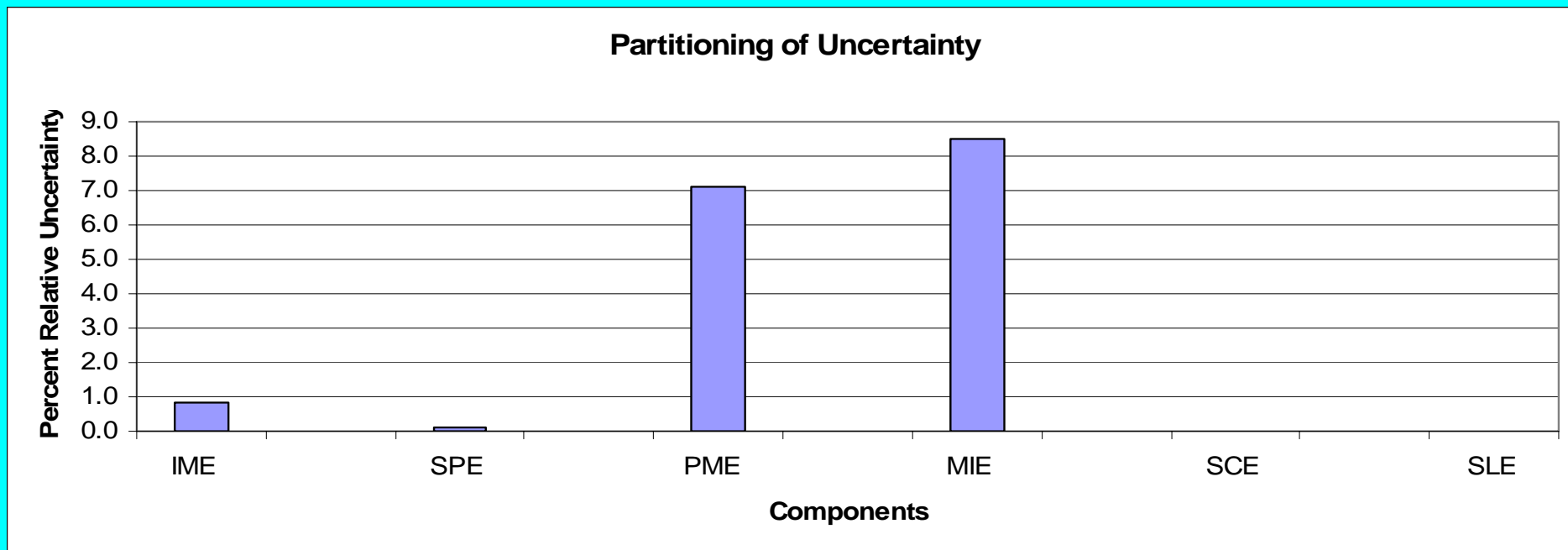
What is the analytical measurement result?   **10**

What are the analytical measurement units?   **mg/L**

If the sample measurement is   10   mg/L   ,
then the uncertainty interval is   7.7   -   12.3 mg/L   at the   95 % Confidence Level (Expanded Uncertainty)

For the above result, if the systematic measurement error (bias) is corrected, and
the corrected measurement is   9.5 mg/L   ,
then the uncertainty interval is   7.3   -   11.7 mg/L   at the   95 % Confidence Level (Expanded Uncertainty)

Perchlorates

- Clean Water Analytical Measurement Variability
  - $RSD_{95\%}$ 11.6%
  - Independent of the Matrix
  - Laboratory Control Sample
- Real World Matrix Analytical Measurement Variability
  - $RSD_{95\%}$ 26.9%
  - Affected by Matrix Interferences
- QC-based Nested Approach
  - Matrix Interference Effect

$$MIE = (26.9\%^2 - 11.6\%^2)^{1/2}$$

$$MIE = 24.3\%$$

# Chemical Oxygen Demand PMBS Pilot Study

- ◆ Initial MQOs Based on Manufacture's Suggestion
  - ◆ Precision: +/- 10% Relative Standard Deviation
  - ◆ Accuracy: 90-110% Recovery of Spiked Samples
- ◆ Recoveries For Both Methods in Reagent Water Acceptable
- ◆ Recoveries For Both Methods in Matrices of Interest Unacceptable

- ◆ Initial MQOs Unachievable
- ◆ New Accuracy (Bias) MQO
  - ◆ Accuracy: 80-120% Recovery of Spiked Samples

Matrix Effects

- ◆ Modified MQOs
  - ◆ Precision: +/-20% Relative Standard Deviation
  - ◆ Accuracy: 80-120% Recovery of Spiked Samples
  - ◆ Objectives Based on Wastewater and Other Regulatory Programs
- ◆ Method 8000 (Approved)
  - ◆ 4 of 8 Labs Achieved MQOs in Wastewater Matrix
- ◆ Method 10125 (New)
 - ◆ 3 of 8 Labs Achieved MQOs in Wastewater Matrix

Cyclic Data

- ◆ Seasonal/Diurnal Data
- ◆ Long-Term Study Data Distribution is Sine Wave
- ◆ Analyze U-shaped Distributions as Two Separate Distributions
- ◆ Separate "Wet Season" Data from "Dry Season" Data
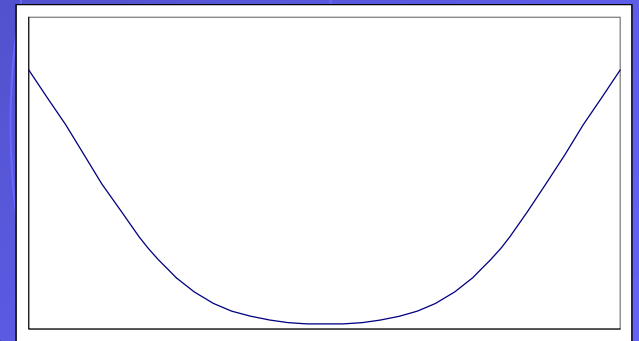
Data Below the Quantification Limit

- ◆ A Single Test Measurement Below the Quantification Limit Cannot be Used to Make a Decision
    - ◆ Average of Replicate Measurements Can Be Used to Make Decisions
- ◆ Random Errors Average Out to Zero
    - ◆ Random Errors for Replicate Measurements Cancel One Another Out
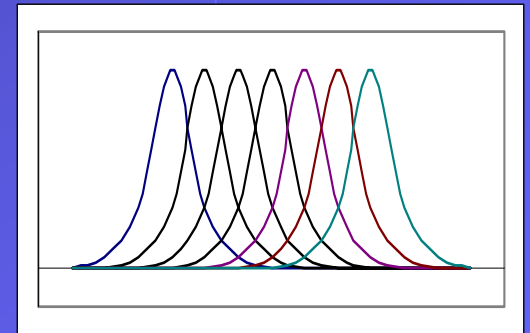
Data Below the Detection Limit

- A Single Non Detect Cannot be Used to Make a Decision

- Measurement Below Detection Limit Are Censored

- Environmental Data is Usually Observed to be Positively Skewed

- Maximum Uncertainty Associated with the Average Measurement Can be Modeled by an Exponential Distribution
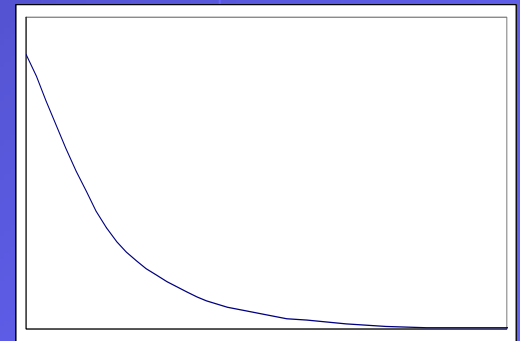
Example of Estimating Average Concentration From Censored Data

- ◆ Mean Can be Calculated From 99% Confidence Level Associated With the Method Detection Limit (MDL)

- ◆ MDL = 10 ppb

- ◆ Mean = 10 ppb/[ln(1-0.99)]

- ◆ Mean = 2 ppb

$$\mu = \frac{-X_q}{\ln(1-q)}$$

Summary

- Water Monitoring Decision-Making Requires Managing Data and Decision Uncertainty

- Quantitative and Qualitative Data Requires Different Approaches to Managing Decision Uncertainty

- Estimation of Data and Decision Uncertainties Enables the Decision-Maker to Compare Data and Make Quality Decisions

Further Information:

Bill Ingersoll

NAVSEA 04XQ (Labs)

843-764-7337

DSN: 794-7337

ingersollws@navsea.navy.mil

www.navylabs.navy.mil